

W/1439

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2003-015938

(43)Date of publication of application : 17.01.2003

(51)Int.Cl.

G06F 12/00

G06F 13/00

(21)Application number : 2001-196738

(71)Applicant : HITACHI LTD

(22)Date of filing : 28.06.2001

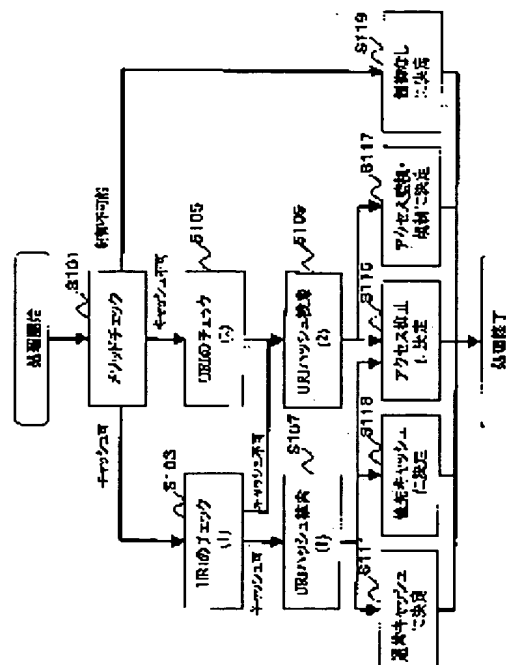
(72)Inventor : AOKI HIDEO
NISHIKADO TAKASHI
YOKOTA DAISUKE
TAKAHASHI YASUHIRO
NODA FUMIO
TAKESHIMA YOSHIAKI

(54) CONGESTION CONTROL METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To avoid or alleviate fault (congestion) that occurs in Web communication relay devices principally involving a large scale Web communication relay device.

SOLUTION: A HTTP module 41 or a caching module 42 included in Web communication relay software 31 makes a method check (S101) regarding if caching is ready or not ready, or control is disabled, and, if the caching is ready, makes a check (1) of a URI (universal resource identifier) (S103). In the step S103, if the caching is ready, URI hash retrieval (1) (S107) is performed to determine any processing of normal caching (S111), priority caching (S113) and access suppression (S115). In the step S101, if the caching is not ready, a check (2) (S105) is made of the URI, and URI hash retrieval (2) (S109) is performed to determine either processing of the access suppression (S115) or an access monitor/restriction (S117). In the step S101, if the control is disabled, a step (S119) without exercising the control is determined.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

W1439

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2003-15938

(P2003-15938A)

(43) 公開日 平成15年1月17日 (2003.1.17)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード* (参考)
G 0 6 F 12/00	5 4 6	G 0 6 F 12/00	5 4 6 K 5 B 0 8 2
13/00	3 5 3	13/00	3 5 3 A 5 B 0 8 9
	5 2 0		5 2 0 C

審査請求 未請求 請求項の数21 O L (全 22 頁)

(21) 出願番号 特願2001-196738 (P2001-196738)

(22) 出願日 平成13年6月28日 (2001.6.28)

(71) 出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72) 発明者 青木 英郎

神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(72) 発明者 西門 隆

神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内

(74) 代理人 100107010

弁理士 橋爪 健

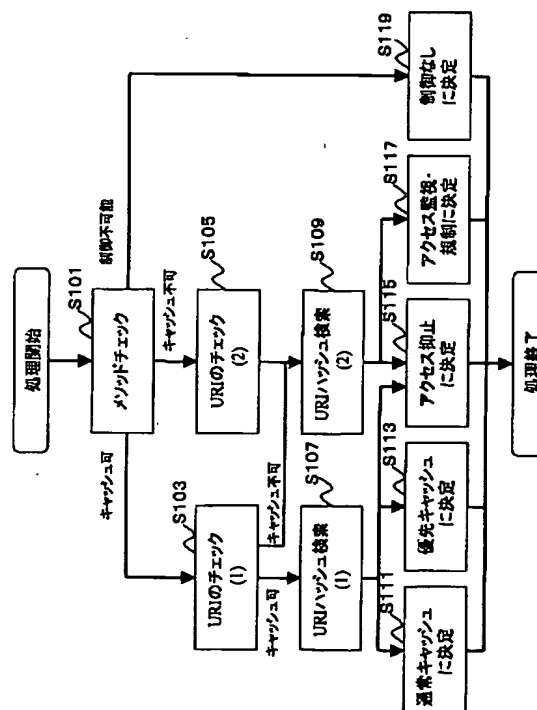
最終頁に続く

(54) 【発明の名称】 輻輳制御方法

(57) 【要約】 (修正有)

【課題】 大規模Web通信中継装置を主としたWeb通信中継装置に発生する障害(輻輳)を回避、軽減する。

【解決手段】 Web通信中継ソフトウェア31に含まれるHTTPモジュール41又はキャッシングモジュール42は、キャッシュ可又は不可か制御不可能かに関してメソッドチェックを行いS101、キャッシュ可であれば、URIのチェック(1)を行うS103。ステップS103でキャッシュ可であれば、URIハッシュ検索(1)S107を行い、通常キャッシュS111、優先キャッシュS113、アクセス抑止S115のいずれかの処理の決定を行う。ステップS101でキャッシュ不可であれば、URIのチェック(2)S105、URIハッシュ検索(2)S109を行い、アクセス抑止S115、アクセス監視・規制S117のいずれかの処理の決定を行う。ステップS101で制御不可能である場合は、制御なしS119に決定する。



【特許請求の範囲】

【請求項 1】クライアント端末とサーバ装置の間に設けられた通信中継装置における輻輳制御方法であって、クライアント端末からの要求に従い、キャッシュヒットした際は記憶されている要求内容をクライアント端末に送信し、キャッシュヒットしなかった場合又はキャッシュを利用しない要求の場合、要求内容をサーバ装置から受信してクライアント端末に送信するための前記輻輳制御方法において、

入力されたクライアント端末からの要求に基づき、キャッシュ可又はキャッシュ不可かを判断するメソッドチェックステップと、

前記メソッドチェックステップによりキャッシュ可であると判断された場合、クライアント端末からの要求に含まれるURIのチェックを行いキャッシュ可又は不可かを判断する第1URIチェックステップと、

前記第1URIチェックステップによる判断に従い、キャッシュ可であれば、URIハッシュの検索を行い、通常キャッシュ、優先キャッシュ又はアクセス抑止のいずれかの処理が行われるか決定する第1URIハッシュ検索ステップと、

前記第1URIハッシュ検索ステップによる決定に従い、通常キャッシュ処理、優先キャッシュ処理又はアクセス抑止処理のいずれかの処理を行うステップとを含む輻輳制御方法。

【請求項 2】クライアント端末とサーバ装置の間に設けられた通信中継装置における輻輳制御方法であって、クライアント端末からの要求に従い、キャッシュヒットした際は記憶されている要求内容をクライアント端末に送信し、キャッシュヒットしなかった場合又はキャッシュを利用しない要求の場合、要求内容をサーバ装置から受信してクライアント端末に送信するための前記輻輳制御方法において、

入力されたクライアント端末からの要求に基づき、キャッシュ可又はキャッシュ不可かを判断するメソッドチェックステップと、

前記メソッドチェックステップによりキャッシュ不可であると判断された場合、クライアント端末からの要求に含まれるURIのチェックを行う第2URIチェックステップと、

URIハッシュの検索を行い、アクセス抑止又はアクセス監視・規制のいずれかの処理が行われるか決定する第2URIハッシュ検索ステップと、

前記第2URIハッシュ検索ステップによる決定に従い、アクセス抑止処理又はアクセス監視・規制処理のいずれかの処理を行うステップとを含む輻輳制御方法。

【請求項 3】クライアント端末とサーバ装置の間に設けられた通信中継装置における輻輳制御方法であって、クライアント端末からの要求に従い、キャッシュヒットした際は記憶されている要求内容をクライアント端末に送

信し、キャッシュヒットしなかった場合又はキャッシュを利用しない要求の場合、要求内容をサーバ装置から受信してクライアント端末に送信するための前記輻輳制御方法において、

入力されたクライアント端末からの要求に基づき、キャッシュ可又はキャッシュ不可かを判断するメソッドチェックステップと、

前記メソッドチェックステップによりキャッシュ可であると判断された場合、クライアント端末からの要求に含まれるURIのチェックを行いキャッシュ可又は不可かを判断する第1URIチェックステップと、

前記第1URIチェックステップによる判断に従い、キャッシュ可であれば、URIハッシュの検索を行い、通常キャッシュ又は優先キャッシュのいずれかの処理が行われるか決定する第1URIハッシュ検索ステップと、前記メソッドチェックステップによりキャッシュ不可であると判断された場合、クライアント端末からの要求に含まれるURIのチェックを行う第2URIチェックステップと、

前記第1URIチェックステップによる判断によりキャッシュ不可である場合又は前記第2URIチェックステップを経て移行され、URIハッシュの検索を行い、アクセス抑止又はアクセス監視・規制のいずれかの処理が行われるか決定する第2URIハッシュ検索ステップと、

前記第1又は第2URIハッシュ検索ステップによる決定に従い、通常キャッシュ処理、優先キャッシュ処理、アクセス抑止処理、アクセス監視・規制処理のいずれかの処理を行うステップとを含む輻輳制御方法。

【請求項 4】前記メソッドチェックステップにより制御不可能であると判断された場合、制御なしの処理を行うステップをさらに含む請求項 1 乃至 3 のいずれかに記載の輻輳制御方法。

【請求項 5】前記メソッドチェックステップは、クライアント端末からの要求に含まれるリクエストメソッドを取得するステップと、

取得されたリクエストメソッドに従い、キャッシュ可又はキャッシュ不可又は制御不可能を決定するステップとを含む請求項 1 乃至 3 のいずれかに記載の輻輳制御方法。

【請求項 6】前記第1URIチェックステップは、クライアント端末からの要求に含まれるスキームがHTTP等の第1のスキームであることを確認するステップと、

URIに付いているクエリを外すことによりURIを加工するステップと、

クライアント端末からの要求に含まれるパスをチェックし、所定の文字を含んでいるかどうか検査するステップと、

所定の文字を含む場合、URIを加工した文字列をメモ

りに確保し、キャッシュ不可の処理へ移行するステップと、
所定の文字を含まない場合は、キャッシュ可の処理へ移行するステップとを含む請求項 1 又は 2 に記載の輻輳制御方法。

【請求項 7】前記第 2 URI チェックステップは、クライアント端末からの要求に含まれるスキームが HTTP 等の第 1 のスキームであることを確認するステップと、
URI に付いているクエリを外すことにより URI を加工するステップとを含む請求項 2 又は 3 に記載の輻輳制御方法。

【請求項 8】前記第 1 又は第 2 URI ハッシュ検索ステップは、
URI ハッシュを検索するステップと、
URI エントリにヒットした場合は、URI エントリに含まれる制御タイプを参照し、そこに記述されている制御に決定するステップと URI エントリにヒットしなかった場合は、予め定められた処理に決定するステップを含む請求項 1 乃至 3 のいずれかに記載の輻輳制御方法。

【請求項 9】URI エントリは、URI と、制御タイプ、アクセスカウンタ、キャッシュエントリ情報を含むことを特徴とする請求項 8 に記載の輻輳制御方法。

【請求項 10】キャッシュエントリは、URI エントリ情報と、各種制御タイプを示すキャッシュエントリタイプと、主として利用しているセッションを示すオーナーセッション情報と、キャッシュデータ情報を含むことを特徴とする請求項 9 に記載の輻輳制御方法。

【請求項 11】制御タイプ又はキャッシュエントリタイプは、通常キャッシュ、優先キャッシュ、アクセス抑止、アクセス監視・規制のいずれかであることを特徴とする請求項 10 に記載の輻輳制御方法。

【請求項 12】前記通常キャッシュ処理又は前記優先キャッシュ処理は、
キャッシュヒットしたか否かを判定するステップと、
キャッシュは有効期限内か否かを判定するステップと、
キャッシュが有効期限内である場合、キャッシュを取出し、クライアント端末へキャッシュを転送するステップと、
キャッシュが有効期限内でない場合、サーバ装置へクライアント端末からの要求を転送し、サーバ装置から受信した応答をクライアント端末へ転送し、さらに、通常キャッシング又は優先キャッシングを行うステップとを含む請求項 1 乃至 11 のいずれかに記載の輻輳制御方法。

【請求項 13】キャッシュがヒットしなかった場合、通常又は優先 URI エントリの作成、通常又は優先キャッシュエントリの作成、URI ハッシュへ登録、キャッシュエントリにオーナーを設定のいずれか又は複数を含むキャッシング前処理を行うステップと、
サーバ装置へクライアント端末からの要求を転送し、サ

ーバ装置から受信した応答をクライアント端末へ転送し、さらに、通常キャッシングを行うステップとを含むことを特徴とする請求項 12 に記載の輻輳制御方法。

【請求項 14】前記キャッシュがヒットした場合、サーバ装置から応答が戻る前に、同一コンテンツに対して複数の要求が重複した場合、最初の要求だけサーバ装置へ中継し、残りの要求は一旦処理待ち状態とし、最初の要求に対する応答が戻ってから処理を再開する集約処理をさらに含むことを特徴とする請求項 12 に記載の輻輳制御方法。

【請求項 15】キャッシュ領域に空きが無くなる又は少なくなったときにアクセス回数が低い領域を消去して再利用するためのガベージコレクション処理の際に、優先キャッシュエントリは対象外とし、さらに優先キャッシュフラグを参照して、優先キャッシュとして使用されているデータ領域も処理の対象外とすることを特徴とする請求項 1 乃至 14 のいずれかに記載の輻輳制御方法。

【請求項 16】前記アクセス抑止処理は、
要求されたコンテンツに関する抑止メッセージを取得するステップと、
抑止メッセージよりコンテンツを作成するステップと、
クライアント端末へコンテンツを転送するステップとを含むことを特徴とする請求項 1 乃至 15 のいずれかに記載の輻輳制御方法。

【請求項 17】前記アクセス監視・規制処理は、
規制を行うか否かを判定する規制判定ステップと、
規制を行う場合、規制メッセージを取得し、規制メッセージよりコンテンツを作成して、クライアント端末へコンテンツを転送するステップと、
規制を行わない場合、サーバ装置へ要求を転送し、サーバ装置から受信した応答をクライアント端末へ転送するステップとを含むことを特徴とする請求項 1 乃至 16 のいずれかに記載の輻輳制御方法。

【請求項 18】URI エントリを作成していなければ作成し、それを URI ハッシュに登録する処理、及び／又は、URI エントリ中のアクセスカウンタを 1 つ増やす処理を実行する監視・規制前処理をさらに含むことを特徴とする請求項 17 に記載の輻輳制御方法。

【請求項 19】前記規制判定ステップは、
① URI エントリのアクセスカウンタ、② URI エントリの規制実行フラグ、③ 装置に設定された規制開始接続数、④ 装置に設定された規制終了接続数を参照して、
②の規制実行フラグが設定されていない場合、①>③ならば規制を行うとし、そうでなければ規制を行わないとし、一方、②の規制実行フラグが設定されている場合、①>④ならば規制を行うとし、そうでなければ規制を行わないとすることを特徴とする請求項 17 に記載の輻輳制御方法。

【請求項 20】URI エントリ中のアクセスカウンタを 1 つ減らし、アクセスカウンタが 0 ならば、URI ハッ

シュから削除し、URI エントリを削除する監視・規制後処理をさらに含むことを特徴とする請求項 17 に記載の輻輳制御方法。

【請求項 21】前記制御なし処理は、サーバ装置へ要求を転送するステップと、サーバ装置から応答を受信するステップと、クライアント端末へ応答を転送するステップとを含むことを特徴とする請求項 1 乃至 20 のいずれかに記載の輻輳制御方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、輻輳制御方法に係り、特に、クライアントとサーバ間のデータ通信における輻輳制御方法に関する。

【0002】

【従来の技術】インターネットが急速に普及し、チケット予約や銀行・証券取引等、従来営業店窓口で行われていたサービスもインターネットを介して提供されるようになってきた。また、通信技術も進歩し、家庭やオフィスからだけでなく携帯電話などからでもサービスを受ける環境が整ってきた。

【0003】しかし、各クライアントが直接サーバにサービス要求をする従来の方式では、クライアントの要求がサーバに集中する。その結果、数多くの利用者からの要求に通信回線容量や処理するサーバ装置の能力が対応できず、時間帯やアクセス先によっては、利用者がいくら要求をしてもなかなか応答が返ってこないという場合が発生するようになってきた。

【0004】この点に関しては、一般に、Web 通信中継装置を設置することで改善できる。クライアントからの要求は、Web 通信中継装置と呼ばれる中継処理装置を経由して処理する。Web 通信中継装置では、過去の通信内容の一部を記憶（キャッシュという）する記憶部（キャッシングモジュール）を設置することにより、記憶した通信に関しては、サーバへ接続することなく要求を処理できる。プロキシ装置は、その有効性から多数の通信システムで導入されている。

【0005】

【発明が解決しようとする課題】現在では、1000 万人規模のユーザを対象とする通信システムに大規模 Web 通信中継装置が設置されるようになってきた。このような通信システムでは、大規模 Web 通信中継装置に対する要求が過剰に集中する。要求の集中が起こると、通信路の混雑や装置の負荷が上昇するため、ユーザは通信システムが提供するサービスを利用しにくくなる。また、過度な集中が継続すると大規模 Web 通信中継装置に障害（輻輳）が発生し動作を停止する可能性がある。大規模 Web 通信中継装置の動作が停止すると通信システムとして機能しなくなり、通信システムを利用する全ユーザがサービスを利用できなくなる。

【0006】本発明では、以上の点を鑑み、大規模 Web 通信中継装置を主とした Web 通信中継装置に発生する障害を回避、軽減することを目的とする。

【0007】

【課題を解決するための手段】上記の課題を解決するため、本発明では、例えば、Web 通信中継装置がユーザの要求する通信内容ごとに要求処理の制御を変更する機構を設ける。Web 通信中継装置の負荷が上昇するのを避けるために、実施する制御を判定する機構は、新たな処理部を設けるのではなく、Web 通信中継装置に従来から存在する記憶部（キャッシングモジュール）を利用する。

【0008】Web 通信中継装置は、現在装置が行っている通信内容と装置の状態を検証したうえで、新たに到着した要求に対する処理方針を決定する。これにより、省略できる通信を検出でき、通信量の削減することで、Web 通信中継装置が輻輳状態に陥ることを回避する。また、輻輳の原因となる通信処理を通常の処理とは異なる方法で処理する。例えば、輻輳の原因となる通信を優先的に記憶する。これにより Web 通信中継装置が輻輳状態に陥るのを回避する。

【0009】Web 通信中継装置は、要求に対する制御を決定する過程で、輻輳状態に近づいている通信を検出する。輻輳状態に近い通信に関しては、輻輳が改善されるまで処理をせず、ユーザに対して通信を控えるようアナウンスする。これにより、輻輳状態の悪化を防止し、Web 通信中継装置が障害で利用不能となるのを回避する。

【0010】本発明の大規模 Web 通信中継装置では、ユーザが送信する要求を負荷分散装置で受信する。負荷分散装置は、大規模 Web 通信中継装置内に設置した複数の Web 通信中継装置に負荷を分散して送信する。Web 通信中継装置では、キャッシングモジュールがユーザの要求を判断し、Web 通信中継装置が行う制御を決定する。輻輳制御の種類としては、要求集約、優先キャッシュ、アクセス抑止、アクセス監視および規制の 4 つがある。

【0011】複数のユーザがキャッシュ可能なサービスの要求を行い、かつ、Web 通信中継装置がサーバに通信する必要がある場合、Web 通信中継装置は、サーバに対しては複数の要求が 1 つの要求に集約して通信する。これにより、Web 通信中継装置とサーバ間の通信量が削減する。また、サーバに対する要求が減少するため、サーバの負荷が上昇するのを防止する。Web 通信中継装置がサーバに対して行う輻輳を回避し、負荷も軽減するため、Web 通信中継装置は、ユーザに迅速に回答できる。

【0012】本発明の Web 通信中継装置は、特定のサービスを優先的にキャッシングする方法を提供する。この方法を利用して、通信システム提供者は輻輳の原因と

なるサービスを優先的にキャッシングするよう登録できる。従来の方法では、キャッシングするデータに優先、非優先の区別がないため、重要なデータをWeb通信中継装置内で保持しつづけることができなかった。しかし、この方法により、Web通信中継装置は、ユーザが必要とするデータを保持できる。ユーザは、より多くのサービスを高速に受けることができる。

【0013】本発明のWeb通信中継装置は、特定のサービスをユーザが利用するのを抑止する方法を提供する。この方法を利用して、通信システム提供者は、輻輳の原因となり通信システムが提供すべきではないサービスをユーザに提供しないことができる。提供しないサービスに関する要求には、Web通信中継装置がユーザに対して提供できない旨の返答を行い要求処理が終了する。

【0014】従来では、抑止機構を持たないため、Web通信中継装置は不要なサーバとの通信に資源を浪費していた。また、ユーザは、要求したサービスが提供されないまま待ち続けたり、何度も通信を試みる可能性があった。本発明の方法を利用することで、サービスが提供されないことが迅速に把握できる。

【0015】本発明のWeb通信中継装置は、サービス接続数が確認できる要求について、その接続数を計測する。要求集約が不可能な要求にもかかわらず、同一の要求が非常に多い場合は、一定数以上の要求を接続不可として処理できる。この際、Web通信中継装置は、ユーザに対して要求結果として接続不可の理由を通知する。これにより、サーバの負荷を軽減するとともにWeb通信中継装置の輻輳を回避する。また、ユーザは、要求したサービスが提供されないまま待つことなく、サービスの提供状況を把握できる。

【0016】本発明の第1の解決手段によると、クライアント端末とサーバ装置の間に設けられた通信中継装置における輻輳制御方法であって、クライアント端末からの要求に従い、キャッシュヒットした際は記憶されている要求内容をクライアント端末に送信し、キャッシュヒットしなかった場合又はキャッシュを利用しない要求の場合、要求内容をサーバ装置から受信してクライアント端末に送信するための前記輻輳制御方法において、入力されたクライアント端末からの要求に基づき、キャッシュ可又はキャッシュ不可かを判断するメソッドチェックステップと、前記メソッドチェックステップによりキャッシュ可であると判断された場合、クライアント端末からの要求に含まれるURIのチェックを行いキャッシュ可又は不可かを判断する第1URIチェックステップと、前記第1URIチェックステップによる判断に従い、キャッシュ可であれば、URIハッシュの検索を行い、通常キャッシュ、優先キャッシュ又はアクセス抑止のいずれかの処理が行われるか決定する第1URIハッシュ検索ステップと、前記第1URIハッシュ検索ステ

ップによる決定に従い、通常キャッシュ処理、優先キャッシュ処理又はアクセス抑止処理のいずれかの処理を行うステップとを含む輻輳制御方法を提供する。

【0017】本発明の第2の解決手段によると、クライアント端末とサーバ装置の間に設けられた通信中継装置における輻輳制御方法であって、クライアント端末からの要求に従い、キャッシュヒットした際は記憶されている要求内容をクライアント端末に送信し、キャッシュヒットしなかった場合又はキャッシュを利用しない要求の場合、要求内容をサーバ装置から受信してクライアント端末に送信するための前記輻輳制御方法において、入力されたクライアント端末からの要求に基づき、キャッシュ可又はキャッシュ不可かを判断するメソッドチェックステップと、前記メソッドチェックステップによりキャッシュ可であると判断された場合、クライアント端末からの要求に含まれるURIのチェックを行う第2URIチェックステップと、URIハッシュの検索を行い、アクセス抑止又はアクセス監視・規制のいずれかの処理が行われるか決定する第2URIハッシュ検索ステップと、前記第2URIハッシュ検索ステップによる決定に従い、アクセス抑止処理又はアクセス監視・規制処理のいずれかの処理を行うステップとを含む輻輳制御方法を提供する。

【0018】本発明の第3の解決手段によると、クライアント端末とサーバ装置の間に設けられた通信中継装置における輻輳制御方法であって、クライアント端末からの要求に従い、キャッシュヒットした際は記憶されている要求内容をクライアント端末に送信し、キャッシュヒットしなかった場合又はキャッシュを利用しない要求の場合、要求内容をサーバ装置から受信してクライアント端末に送信するための前記輻輳制御方法において、入力されたクライアント端末からの要求に基づき、キャッシュ可又はキャッシュ不可かを判断するメソッドチェックステップと、前記メソッドチェックステップによりキャッシュ可であると判断された場合、クライアント端末からの要求に含まれるURIのチェックを行いキャッシュ可又は不可かを判断する第1URIチェックステップと、前記第1URIチェックステップによる判断に従い、キャッシュ可であれば、URIハッシュの検索を行い、通常キャッシュ又は優先キャッシュのいずれかの処理が行われるか決定する第1URIハッシュ検索ステップと、前記メソッドチェックステップによりキャッシュ不可であると判断された場合、クライアント端末からの要求に含まれるURIのチェックを行う第2URIチェックステップと、前記第1URIチェックステップによる判断によりキャッシュ不可である場合又は前記第2URIチェックステップを経て移行され、URIハッシュの検索を行い、アクセス抑止又はアクセス監視・規制のいずれかの処理が行われるか決定する第2URIハッシュ検索ステップと、前記第1又は第2URIハッシュ検

索ステップによる決定に従い、通常キャッシュ処理、優先キャッシュ処理、アクセス抑止処理、アクセス監視・規制処理のいずれかの処理を行うステップとを含む輻輳制御方法を提供する。

【0019】

【発明の実施の形態】 A. ハード構成と概略構成

図1に、通信システムの構成図を示す。通信システム100は、例えば、クライアント端末1-1~1-iと、大規模Web通信中継装置2と、Web通信装置3-1~3-jとを含む。クライアント端末1-1~1-iは、例えば、Webコンテンツの取得（アクセス）要求、データ送受信、画面表示等を行うユーザ用（端末）装置である。Web通信装置3-1~3-jは、例えば、クライアント端末1-1~1-iからの要求に応じてWebコンテンツを送出する装置であって、ディスク容量を上限としてコンテンツデータを格納している。なお、本実施の形態では、主に大規模なWeb通信装置について説明しているが、これに限らず、本発明は、適宜の規模の通信装置に適用することができる。

【0020】大規模Web通信中継装置2は、例えば、データ中継装置21と、認証装置22と、システム管理装置23と、レイヤ7負荷分散装置24と、Web通信中継装置25-1~25-nと、アクセス規制装置26とを備える。

【0021】データ中継装置21は、例えば、大規模Web通信中継装置2内外の各装置、Web通信装置3-1~3-j及びクライアント端末1-1~1-iとの間でデータ送受信を中継する装置であって、送受信データパケットのヘッダーに記された宛先装置へパケットを転送する。認証装置22は、大規模Web通信中継装置2が提供する中継サービスが、予め登録された（正規の）クライアント端末1-1~1-iと正規のWeb通信装置3-1~3-j間のものであることを認証するための装置であって、ユーザ情報やサービス情報を登録したデータベースを備える。システム管理装置23は、例えば、大規模Web通信中継装置2内の各装置を集中的に管理するための装置である。

【0022】レイヤ7負荷分散装置24は、例えば、複数のクライアント端末1-1~1-iからのWebコンテンツ取得要求を、複数のWeb通信中継装置25-1~25-nに振分け、中継処理の負荷を分散する装置であって、送受信パケットヘッダーや、Webコンテンツ要求・送信ヘッダーに記された宛先等の記述を解釈して振分け経路を定め、受信したデータパケットをその経路へ送信する。Web通信中継装置25-1~25-nは、例えば、クライアント端末1-1~1-iからのWebコンテンツ取得要求を受け付け、Web通信装置3-1~3-jに要求を中継したり、クライアント端末1-1~1-iとWeb通信装置3-1~3-j間のデータ送受信を中継する装置である。アクセス規制装置26

は、例えば、多量のWebコンテンツ取得要求が短時間に集中した場合に、大規模Web通信中継装置2の処理限界を超えないよう、アクセス量（取得要求量）を規制するための装置である。

【0023】図2に、Web通信中継装置の構成図を示す。Web通信中継装置25-1は、例えば、主記憶装置30と、この主記憶装置30に記憶されたWeb通信中継ソフトウェア31と、入出力装置32と、プロセッサ33と、レイヤ7負荷分散装置24とデータパケットの送受信を行う通信装置34と、二次記憶装置35と、これらの各部材を相互の接続するバス36とを備える。

【0024】二次記憶装置35は、例えば、磁気ディスクなどの装置（不揮発性記憶装置）であって、プログラム及び各種の設定ファイルを記憶する。二次記憶装置35は、以下のような役割を担う。

（1）Web通信中継装置25-1のOS（オペレーティングシステム）データを格納し、装置起動時にそのデータが主記憶装置30にロードされる。

（2）OS上で動作する様々なアプリケーションプログラム（データ・設定ファイル）を格納し、アプリケーションの起動時に主記憶装置30にロードされる。

（3）Web通信中継装置25-1が中継するWebコンテンツデータをキャッシュデータとして格納する。

【0025】なお、実施例のWeb通信中継装置では、クライアントからの要求をセッションと呼ばれる処理単位で処理している（1つの要求が1つのセッションに対応する）。セッションがキャッシュを用いる際にキャッシュエントリを確保する。セッションがサーバと通信してキャッシングをするときは、キャッシュエントリを所有している必要がある。本輻輳制御方式は、個々のセッション処理の中で行われる。

【0026】図3は、大規模Web通信中継装置におけるデータの送受信についての説明図である。なお、説明の便宜上、上述した部材と同一部材には同一符号を付し、構成、機能は同様である。また、各Web通信中継装置25-1~25-nは同様の構成であるので、ここでは、Web通信中継装置25-1について説明する。

【0027】Web通信中継ソフトウェア31は、Web通信中継装置25-1に記憶されており、例えば、HTTPモジュール41と、キャッシングモジュール42と、プロキシエンジン43とを含む。ここで、図中のデータ①~④は、①クライアント端末1-1~1-iの要求、②Web通信中継ソフトウェア31の応答、③Web通信中継ソフトウェア31の要求、④Web通信装置3-1~3-jの応答をそれぞれ示す。

【0028】なお、キャッシュヒットのときは、データ①、②のみが通信路を流れる（図4参照）。また、キャッシュミスヒットのときは、データ①、③、④、②の順でデータが通信路を流れる（図5参照）。以下に各処理について概略説明する。

【0029】①クライアントの要求

まず、クライアント端末1-1~1-iからクライアント要求が送信されると、大規模Web通信中継装置2は、それを受信する。大規模Web通信中継装置2のデータ中継装置21では、要求をレイヤ7負荷分散装置24に転送し、レイヤ7負荷分散装置24は、いずれかのWeb通信中継装置（ここでは、Web中継装置25-1）に送る。Web通信中継装置25-1内のWeb通信中継ソフトウェア31では、プロキシエンジン43を経て、HTTPモジュール41とキャッシングモジュール42が、そのクライアント要求を受信する。

【0030】②Web通信中継ソフトウェアの応答

つぎに、クライアント要求を受信したWeb通信中継装置25-1は、通信内容を解析する。ここで、Web中継装置25-1が、クライアント要求に対応しているWebページをキャッシュで記憶している場合、その内容をレイヤ7負荷分散装置24、データ中継装置21を経て、クライアント端末1-1へ送信する。

【0031】③Web通信中継ソフトウェアの応答

一方、Web中継装置25-1が、クライアント要求に対応しているWebページをキャッシュで記憶していない場合、及び／又はクライアント要求にキャッシュ利用不可の指示がある場合、Web中継装置25-1は、クライアント要求に関する情報を、レイヤ7負荷分散装置24、データ中継装置21を経て、Web通信装置3-1~3-jの該当する装置へ送信する。

【0032】④Web通信装置の応答

Web通信装置3-1~3-jは、クライアント要求を受信すると、クライアント端末に、データ中継装置21を介してクライアント要求に対応するWebページを応答する。さらに、データ中継装置21は、レイヤ7負荷分散装置24を経てWeb中継装置25-1に、その通信内容やWebページを転送する。Web中継装置25-1では、そのWebページをキャッシングモジュール42にキャッシュする。

【0033】図4は、キャッシュヒット時の説明図及びシーケンス図である。なお、ここでのシーケンスは、構成図中の点線の矢印に対応している。まず、クライアント端末1-1~1-iからの要求信号（データ①）は、大規模Web通信中継装置2内のデータ中継装置21、レイヤ7負荷分散装置24を経て、Web通信中継ソフトウェア31内のHTTPモジュール41を経て、キャッシングモジュール42に入力され、ここで、キャッシュヒットと判定される。つぎに、Webページは、Web通信中継ソフトウェア31の応答信号（データ②）として、再び、キャッシングモジュール42、HTTPモジュール41、レイヤ7負荷分散装置24、データ中継装置21を経て、クライアント端末1-1~1-iに送信される。

【0034】図5は、キャッシュミスヒット時の説明図

及びシーケンス図である。なお、ここでのシーケンスは、構成図中の点線の矢印に対応している。まず、クライアント端末1-1~1-iからの要求信号（データ①）は、大規模Web通信中継装置2内のデータ中継装置21、レイヤ7負荷分散装置24を経て、Web通信中継ソフトウェア31内のHTTPモジュール41を経て、キャッシングモジュール42に入力され、ここで、キャッシュミスヒットと判定される。つぎに、Web通信中継ソフトウェア31の要求信号（データ③）が、キャッシングモジュール42、HTTPモジュール41、レイヤ7負荷分散装置24、データ中継装置21を経て、Web通信装置3-1~3-jに送信される。さらに、これに回答してWeb通信装置3-1~3-jのWebページを含む応答信号（データ④）は、大規模Web通信中継装置2内のデータ中継装置21、レイヤ7負荷分散装置24を経て、Web通信中継ソフトウェア31内のHTTPモジュール41を経て、キャッシングモジュール42に入力される。つぎに、Webページは、Web通信中継ソフトウェア31の応答信号（データ②）として、再び、キャッシングモジュール42、HTTPモジュール41、レイヤ7負荷分散装置24、データ中継装置21を経て、クライアント端末1-1~1-iに送信される。

【0035】図6は、輻輳時の説明図及びシーケンス図である。なお、ここでのシーケンスは、構成図中の点線の矢印に対応している。まず、クライアント端末1-1~1-iからの要求信号（データ①）は、大規模Web通信中継装置2内のデータ中継装置21、レイヤ7負荷分散装置24を経て、Web通信中継ソフトウェア31内のHTTPモジュール41を経て、キャッシングモジュール42に入力される。

【0036】ここで、Web通信中継ソフトウェア31は、例えば、Web通信装置3-1~3-jが輻輳中の場合、規制コンテンツを生成する。ここで、例えば、ある一定以上の要求がひとつのWebページに集中した場合、輻輳と判断される。この規制コンテンツは、データ②に含まれ、クライアント端末1-1~1-iに送信される。その際、規制コンテンツは、Web通信中継ソフトウェア31の応答信号（データ②）として、再び、キャッシングモジュール42、HTTPモジュール41、レイヤ7負荷分散装置24、データ中継装置21を経て、クライアント端末1-1~1-iに送信される。

【0037】図7は、アクセス抑止対象コンテンツが要求された場合の説明図及びシーケンス図である。なお、ここでのシーケンスは、構成図中の点線の矢印に対応している。まず、クライアント端末1-1~1-iからのアクセス抑止対象コンテンツを含む要求信号（データ①）は、大規模Web通信中継装置2内のデータ中継装置21、レイヤ7負荷分散装置24を経て、Web通信中継ソフトウェア31内のHTTPモジュール41を経

て、キャッシングモジュール42に入力される。ここで、Web通信中継ソフトウェア31は、例えば、クライアント端末1-1~1-iからアクセス抑止対象コンテンツが要求されている、抑止アナウンスを生成する。この抑止アナウンスは、データ②に含まれ、クライアント端末1-1~1-iに送信される。その際、抑止アナウンスは、Web通信中継ソフトウェア31の応答信号(データ②)として、再び、キャッシングモジュール42、HTTPモジュール41、レイヤ7負荷分散装置24、データ中継装置21を経て、クライアント端末1-1~1-iに送信される。

【0038】図8は、Web通信中継ソフトウェア31の内部構成についての説明図である。Web通信中継ソフトウェア31は、主記憶装置30に記憶される。Web通信中継ソフトウェア31は、例えば、データ領域50と、セッション処理プロセス45とを含む。セッション処理プロセス45は、例えば、HTTPモジュール41、キャッシングモジュール42、プロキシエンジン43を含む。プロキシエンジン43は、Web通信中継ソフトウェア31の基本機能を実現し、通信データ送受信やセッション管理などを行う。HTTPモジュール41は、HTTPパケットの処理を行う。キャッシングモジュール42は、HTTPのキャッシングと輻輳制御を行う。また、データ領域は、例えば、プロキシエンジン使用領域、HTTPモジュール使用領域、キャッシングモジュール使用領域を含む。さらに、キャッシングモジュール使用領域は、例えば、URIハッシュ、URIエン트리及びURI文字列、アクセス規制/抑止、メッセージ及びフォーマット、キャッシュエン트리領域、キャッシュデータ領域を含む。

【0039】図9は、URIハッシュとURIエン트리によるURI管理についての説明図である。URIハッシュ70は、例えば、Key0~nのURIエントリのリスト70-0~70-nを含む。Key0のURIエントリのリスト70-0は、例えば、URIエン트리71として、URI1(監視・規制)71-1と、URI2(抑止)71-2とを含む。また、Key2のURIエントリのリスト70-2は、例えば、URIエン트리71として、URI3(監視・規制)71-3と、URI4(通常キャッシュ)71-4と、URI5(優先キャッシュ)71-5とを含む。Key nのURIエントリのリスト70-nは、例えば、URIx(通常キャッシュ)71-xを含む。

【0040】ここで、URI4(通常キャッシュ)71-4とURIx(通常キャッシュ)71-xは、通常キャッシュエントリに通常キャッシュデータを含む。また、URI5(優先キャッシュ)71-5は、優先キャッシュエントリに優先キャッシュデータを含む。ここで、URI管理について詳細に説明する。URIエントリは、URIをハッシュ関数にかけて算出したハッシュ

キーを元にして、同一ハッシュキーでリスト化して管理している。この例では、Key0にはURI1(監視・規制)71-1、URI2(抑止)71-2、Key2にはURI3(監視・規制)71-3、URI4(通常キャッシュ)71-4、…のURIエントリがリストされている。また、リスト中の各々のURIは、要求集約で監視規制対象になっていたり、オペレータ指示で抑止対象や優先キャッシュ対象になっていたり、通常キャッシュであったり、渾然一体となってリスト化されている。このうち、例えば、通常キャッシュであるURI4(通常キャッシュ)71-4のエントリは、通常キャッシュエントリの特定のアドレスを示し、同様に、優先キャッシュであるURI5(優先キャッシュ)71-5のエントリは、優先キャッシュエントリの特定のアドレスを示している。

【0041】また、主記憶装置30上のWeb通信中継ソフトウェア内キャッシングモジュール使用領域には、例えば、通常キャッシュエントリ領域、優先キャッシュエントリ領域、キャッシュデータ領域が設けてあり、それぞれ所定の大きさに区切って準備されている。実際のURI4(通常キャッシュ)71-4のコンテンツデータは、URI4(通常キャッシュ)71-4のキャッシュエントリで示されたアドレスの通常キャッシュエントリ領域に格納され、コンテンツの大きさによって必要な量だけ通常キャッシュデータ領域が追加される。同様に、実際のURI5(優先キャッシュ)71-5のコンテンツデータは、URI5(優先キャッシュ)71-5のキャッシュエントリで示されたアドレスの優先キャッシュエントリ領域に格納され、コンテンツの大きさによって必要な量だけ優先キャッシュデータ領域が追加される。

【0042】つぎに、主記憶装置30に記憶されるキャッシングモジュール使用領域について説明する。なお、キャッシングモジュール使用領域は、主記憶装置30以外の適宜の記憶装置内に設けることができる。図10は、キャッシングモジュール使用領域内のデータ構造についての説明図(1)である。キャッシングモジュール使用領域は、例えば、URIエン트리71、キャッシュエン트리72とを含む。ここで、URIエン트리71としては、例えば、URI(例: <http://www.abcde.co.jp/index.html>)、制御タイプ(例: 通常キャッシュ)、アクセスカウンタ(例: 3)、キャッシュエントリ情報を含む。

【0043】キャッシュエン트리72は、例えば、URIエントリ情報(例: <http://www.abcde.co.jp/index.html>に関するURIエントリ)、キャッシュエントリタイプ(例: 通常キャッシュ)、キャッシュの状態(例: キャッシング中)、オーナーセッション(例: セッションNo. n)、応答待ちセッションリスト(例: セッションNo. x, No. y)、キャッシュデータ情報を含む。

む。なお、優先キャッシュエントリと通常キャッシュエントリは、キャッシュエントリタイプの値が異なるが、内容は同様である。また、キャッシングモジュール42内では、優先キャッシュエントリは、運用者の要求がない限り削除されない。但し、通常キャッシュエントリは、資源が不足すると削除される。

【0044】図11は、キャッシングモジュール使用領域内のデータ構造についての説明図(2)である。キャッシングモジュール使用領域は、また、例えば、キャッシュデータ73を含む。キャッシュデータ73は、例えば、キャッシュエントリ情報(例: <http://www.abcde.co.jp/index.html>に関するキャッシュエントリ)、キャッシュデータタイプ(例: 通常キャッシュ)、キャッシュデータを参照しているセッション数(例: 3)、次のキャッシュデータ情報、キャッシュデータを含む。なお、優先キャッシュデータと通常キャッシュデータは、キャッシュデータタイプの値が異なるが、内容は同様である。また、キャッシングモジュール42内では、優先キャッシュデータは、運用者の要求がない限り削除されない。但し、通常キャッシュデータは、資源が不足すると削除される。

【0045】B. 詳細動作

つぎに、キャッシング及び輻輳制御処理を説明する。図12は、キャッシング及び輻輳制御処理について基本的な処理を示すフローチャートである。まず、Web通信中継ソフトウェア31は、キャッシュチェックおよび制御種別判定を行う(S201)。Web通信中継ソフトウェア31は、この判定結果に応じて、例えば、通常キャッシュの制御(S203)、優先キャッシュの制御(S205)、アクセス抑止の制御(S207)、アクセス監視・規制の制御(S209)、制御なし(S211)のいずれか又は複数の制御を行い、処理を終了する。

【0046】以下に、キャッシュチェックおよび輻輳制御処理判定について詳細に説明する。図13は、キャッシュチェックおよび輻輳制御処理判定についてのフローチャートである。まず、Web通信中継ソフトウェア31に含まれるキャッシングモジュール42は、キャッシュ可又は不可か制御不可能かに関してメソッドチェックを行い(S101)、キャッシュ可であれば、URIのチェック(1)を行う(S103)。キャッシングモジュール42は、ステップS103でキャッシュ可であれば、URIハッシュ検索(1)(S107)を行い、通常キャッシュ(S111)、優先キャッシュ(S113)、アクセス抑止(S115)のいずれかの処理の決定を行う。

【0047】また、キャッシングモジュール42は、ステップS101のメソッドチェックにおいてキャッシュ不可であれば、URIのチェック(2)(S105)を行い、さらに、URIハッシュ検索(2)(S109)

を行い、アクセス抑止(S115)、アクセス監視・規制(S117)のいずれかの処理の決定を行う。なお、このURIハッシュ検索(2)は、ステップS103でキャッシュ不可である場合でも行われる。また、HTTPモジュール41又はキャッシングモジュール42は、ステップS101のメソッドチェックで制御不可能である場合は、制御なし(S119)に決定する。ここで、上述の各処理について詳細に説明する。

【0048】(S101のメソッドチェック)メソッドチェックでは、次のような処理が実行される。

①まず、クライアントのHTTP要求のリクエストメソッドを取得する。

②ここで、リクエストメソッドがGET、HEADなら「キャッシュ可」の場合の処理を行う。

③一方、リクエストメソッドがOPTIONS、POST、DELETE、TRACE、CONNECTなら「キャッシュ不可」の場合の処理を行う。

④さらに、リクエストメソッドがRFC2616で定義されていない拡張メソッドである場合は、「制御不可能」の場合の処理を行う。

【0049】図14は、クライアントのHTTP要求についての説明図である。クライアントのHTTP要求80は、例えば、図示のように、リクエストメソッド(ここでは、GET)と、URI(ここでは、<http://www.abcde.co.jp/>)とを含む。この例では、上述の処理①でリクエストメソッドとして「GET」が取得され、「キャッシュ可」の処理に移行する。

【0050】(ステップS103、105のURIチェック(1)(2))

①ここでは、まず、スキームがHTTPであることを確認する。

②つぎに、パスをチェックし、文字「?」を含んでいるかどうか検査する。

③また、パスをチェックし、.cgi,.aspを含んでいるかどうか検査する。

④さらに、クエリ有無を判断して、クエリを外す加工処理を行う。なお、「クエリが有る」とは、「?」がついていることを指す。

⑤②、③で該当するものがあった場合は「URIの加工」をおこなった文字列を共用メモリに確保する。

⑥②、③で該当した場合は、「キャッシュ不可」の場合の処理を行い、一方、②、③で該当しなかった場合は、「キャッシュ可」の場合で処理を進める。

【0051】図15は、URIの加工についての説明図である。URIは、例えば、図示のように、URI(例えば、<http://www.abcde.co.jp/index.html>)のうち、「http:」をスキーム、「//www.abcde.co.jp/index.html」をパスとする。また、URIの加工とは、例えば、図示のように、URI(例えば、<http://www.abcde.co.jp/a.cgi?a1=arg1&a2=arg2>)に付加されているクエリを

外すことである。この例では、クエリは、'?'al=arg1&a2=arg2'であり、これを除いて、http://www.abcde.co.jp/a.cgi とする。

【0052】このように、URIチェックでは、URIハッシュ検索の前に、CGIなど動的に生成されるページのチェックを行う。具体的には、URIで示されたファイルの識別子と、CGI特有のパラメータ（引数）の有無によって判定を行い、CGIなど動的に生成されるページであると判定した場合には、引数の省略処理を行って、簡略化した短いURIを生成する。これにより、さまざまなクエリを持つCGIを、クエリを含めた別々のURIとして処理するのではなく、1つのURIとして処理し、アクセスの監視・規制を実現する。また、URIのチェック（1）では、上記判定と共に、キャッシュ可否の判定を行う。一方、URIのチェック（2）では、予めキャッシュ不可のURIであることが分かっているため、キャッシュ可否判定処理を行わない。

【0053】（ステップS107、109のURIハッシュ検索（1）（2））ステップS107のURIハッシュ検索（1）では、次のような処理を実行する。例えば、メモリ上にキャッシュの領域があり、クライアント端末と通信しているURIはメモリ上に記憶されており、そのメモリテーブルをURIハッシュとすることができる。

- ①URIハッシュを検索する。
- ②URIエントリにヒットした場合は、URIエントリの制御タイプを参照し、そこに記述されている制御に決定する。
- ③URIエントリにヒットしなかった場合は、「通常キャッシュ」処理に決定する。

【0054】ステップS109のURIハッシュ検索（2）では、

- ①「URIの加工」が行われなかった場合は、HTTP要求にあるURIを使ってURIハッシュを検索する。
- ②URIエントリにヒットした場合は、URIエントリの制御フラグを参照し、そこに記述されている制御に決定する。
- ③URIエントリにヒットしなかった場合は、「アクセス監視・規制」処理に決定する。

【0055】（ステップS111の通常キャッシュに決定）図16は、通常キャッシュ処理についてのフローチャートである。

（1）まず、キャッシュヒットを判定する（S111-1）。

（1-1）ヒットしなかった場合は、最初のアクセス要求としてWebサーバへ要求を中継する処理に移り、キャッシングモジュール42は、例えば、①通常URIエントリの作成、②通常キャッシュエントリの作成、③URIハッシュへ登録、④キャッシュエントリにオーナーの設定、等を行う通常キャッシング前処理を行う（S11

1-11）。つぎに、Webサーバへ要求を転送し（S111-12）、Webサーバから応答を受信し（S111-13）、さらに、クライアントへ応答を転送する（S111-14）。つぎに、キャッシングモジュール42は、通常キャッシングを行う（ステップS111-15）。ここで、ステップS111-15の通常キャッシングでは、例えば、次の処理が行われる。

- ①通常キャッシュデータの確保。
- ②Web通信装置の応答をコピー。
- ③ ①、②をデータがあるまで続ける。
- ④応答待ちセッションリストに登録があるか検査する。
- ⑤ ④で応答待ちセッションがある場合は、処理開始指示を送る。
- ⑥オーナーの設定を解除する。

【0056】（1-2）ヒットした場合、要求集約判定に移る（S111-2）。ここでは、キャッシングモジュール42は、例えば、Webサーバへのアクセス負荷を軽減するため、キャッシュできるコンテンツに対して、アクセス要求を一つに集約する機能を備える。この機能は、Webサーバから応答が戻る前に、同一コンテンツに対して複数の要求が重複した場合、最初の要求だけWebサーバへ中継し、残りの要求は一旦処理待ち状態とし、最初の要求に対する応答が戻ってから処理を再開するものである。具体的には、集約処理を以下のように行う。ここで、ステップS111-2の集約判定処理では、①キャッシュの状態は、準備完了となっているかを判定し、準備完了となっているのであれば、Noとする。また、②キャッシュエントリのオーナーは、現在判定処理を行っているセッションかを判定し、他のセッションがオーナーである場合は、Yesとし、そうでなければ、Noとする。

【0057】（2）要求集約判定でYesの場合、Webサーバへ中継した要求の処理が終わるまで、処理を中止する（S111-3）。ここで、ステップS111-3の処理可能となるまで待機する処理では、①キャッシュエントリの応答待ちセッションリストにセッションを登録し、②呼び出されるまで待機する。

【0058】（3）つぎに、ステップS111-3の後、処理が可能になると、又は、ステップS111-2で集約を行わない場合、キャッシュ有効期限判定に移る（S111-4）。ここで、ステップS111-4のキャッシュ有効期限の判定では、キャッシュ中のRFC2616に記述されている通りに有効期限を計算し、期限内であれば、Yesとし、期限切れであれば、Noとする。

【0059】キャッシュ有効期限判定でYesの場合、キャッシュからデータを取り出す（S111-5）。ここで、ステップS111-5のキャッシュ取出しでは、①必要な分だけキャッシュデータをアクセスする。つぎに、ステップS111-5で取り出したデータを、クラ

クライアントへデータを転送する(S111-6)。なお、通常、このケースでは、直前に新鮮なデータがキャッシュされている確率が高い(何故なら、最初の要求がWebサーバからデータを取ってくるのを待っていた)ので、キャッシュが有効期限内にある確率も高い。結果として、Webサーバへのアクセス要求数が減り、アクセス負荷を軽減できる。

【0060】(4)一方、キャッシュ有効期限判定でNoの場合、Webサーバへ要求を転送し(S111-7)、Webサーバから応答を受信し(S111-8)、さらに、クライアントへ応答を転送する(S111-9)。つぎに、キャッシングモジュール42は、通常キャッシングを行う(S111-10)。なお、(1-1)で、最初のアクセスとして処理した要求は、通常キャッシング処理(S111-10)で、処理が中止になっている要求が存在するか検査する。存在した場合は、処理の再開指示を出す。

【0061】なお、ステップS111-1の前に、後述のような監視・規制の処理を追加することもできる。その際、後述のように、まず、監視・規制前処理(S117-1)を追加して、その後に、規制を行うか否か判定する(S117-2)。ここで、規制を行わない場合、ステップS111-1に移行する。一方、規制を行う場合、規制メッセージを取得する処理(S117-3)、規制メッセージよりコンテンツを作成する処理(S117-4)、及び、クライアントへコンテンツを転送する処理(S117-5)を実行する。さらに、ステップS111-6、S111-10、S111-15の後に、監視・規制後処理(S117-9)が追加され、その実行後、処理終了となる。

【0062】図17は、Web通信装置のHTTP応答の説明図である。このWeb通信装置のHTTP応答85は、例えば、ステップS111-8、13に含まれるものである。

【0063】図18は、Webコンテンツデータの管理についての説明図である。まず、Web通信中継ソフトウェア31内のキャッシングモジュール42は、Web通信装置HTTP応答85を受信する。キャッシングモジュール42は、Web通信装置HTTP応答85を、例えば、Web通信装置HTTP応答Part 1を含むキャッシュエントリ55-1と、Web通信装置応答データPart 2を含むキャッシュデータ55-2と、Web通信装置応答データPart nを含むキャッシュデータ55-nとに区分し、キャッシングモジュール使用領域に格納する。

【0064】(ステップS113の優先キャッシュに決定)図19は、優先キャッシュ処理についてのフローチャートである。なお、ここでのフローチャートは、上述したステップS111の通常キャッシュ処理と比較して、通常キャッシュが優先キャッシュとなったことで変

更された処理以外は、同様である。まず、キャッシングモジュール42は、キャッシュはヒットしたか否かを判定し(S113-1)、キャッシュヒットがあれば、集約を行うか否かを判定する(S113-2)。ここで、ステップS113-2の集約判定処理では、①キャッシュの状態は、準備完了となっているかを判定し、準備完了となっているのであれば、Noする。また、②優先キャッシュエントリのオーナーは、現在判定処理を行っているセッションかを判定し、他のセッションが持ち主である場合は、Yesとし、そうでなければ、Noとする。

【0065】つぎに、キャッシングモジュール42は、ステップS113-2で集約を行う場合、処理可能になるまで待つ(S113-3)。ここで、ステップS113-3の処理可能になるまで待つ処理では、①キャッシュエントリの応答待ちセッションリストにセッションを登録し、②呼び出されるまで待機する。つぎに、キャッシングモジュール42は、キャッシュは有効期限内か否かを判定する(S113-4)。ここで、ステップS113-4のキャッシュ有効期限の判定では、RFC2616に記述されている通りに有効期限を計算し、期限内であれば、Yesとし、期限切れであれば、Noとする。ここで、キャッシュが有効期限内であれば、キャッシュを取出す(S113-5)。ここで、ステップS113-5のキャッシュ取出しでは、①必要な分だけキャッシュデータをアクセスする。つぎに、キャッシングモジュール42は、クライアントへキャッシュを転送する(S113-6)。

【0066】また、キャッシングモジュール42は、ステップS113-4でキャッシュが有効期限内でない場合、Webサーバへ要求を転送し(S113-7)、Webサーバから応答を受信し(S113-8)、さらに、クライアントへ応答を転送する(S113-9)。つぎに、キャッシングモジュール42は、優先キャッシングを行う(S113-10)。ここで、ステップS113-10の優先キャッシングでは、①優先キャッシュデータの確保、②Web通信装置の応答をコピー、③①、②をデータがあるまで続ける、④応答待ちセッションリストに登録があるか検査する、⑤④で応答待ちセッションがある場合は、処理開始指示を送る、⑥オーナーの設定を解除する、等を行う。

【0067】また、キャッシングモジュール42は、ステップS113-1でキャッシュがヒットしなかった場合、通常キャッシング前処理を行う(S113-11)。ここで、ステップS113-11の優先キャッシング前処理としては、例えば、①優先URIエントリの作成、②優先キャッシュエントリの作成、③URIハッシュへ登録、④優先キャッシュエントリにオーナーを設定することが挙げられる。

【0068】また、キャッシングモジュール42は、ステップS113-11の後、Webサーバへ要求を転送

し(S113-12)、Webサーバから応答を受信し(S113-13)、さらに、クライアントへ応答を転送する(S113-14)。つぎに、キャッシングモジュール42は、優先キャッシングを行う(S113-15)。なお、ステップS113-15での優先キャッシングは、ステップS113-10での処理と同様である。

【0069】なお、ステップS113-1の前に、後述のような監視・規制の処理を追加することもできる。その際、後述のように、まず、監視・規制前処理(S117-1)を実行し、その後に、規制を行うか否かを判定する(S117-2)。ここで、規制を行わない場合、ステップS111-1に移行する。一方、規制を行う場合、規制メッセージを取得する処理(S117-3)、規制メッセージよりコンテンツを作成する処理(S117-4)、及び、クライアントへコンテンツを転送する処理(S117-5)を実行する。さらに、ステップS113-6、S113-10、S113-15の後に、監視・規制後処理(S117-9)が追加され、その実行後、処理終了となる。

【0070】(ステップS115のアクセス抑止に決定)アクセス抑止コンテンツの作成方法としては、例えば、アクセス抑止メッセージとアクセス抑止コンテンツのフォーマットを用意して動的に生成する。例えば、上に、アクセス抑止メッセージ群とアクセス抑止コンテンツのフォーマットを格納しておく。キャッシングモジュール42は、状態に応じてふさわしいメッセージを取り出し、アクセス抑止コンテンツフォーマットに挿入することでアクセス抑止コンテンツを生成する。例えば、以下の図に示すように、抑止時の一般的なメッセージ93「要求されたコンテンツは利用できません」を使って、アクセス抑止メッセージフォーマットに挿入し、アクセス抑止コンテンツを生成している。

【0071】図20は、アクセス抑止処理についてのフローチャートである。まず、キャッシングモジュール42は、抑止メッセージを取得し(S115-1)、抑止メッセージよりコンテンツを作成して(S115-2)、クライアントへコンテンツを転送する(S115-3)。

【0072】図21は、抑止メッセージ、抑止コンテンツの説明図である。抑止メッセージ91は、例えば、「要求されたコンテンツはご利用できません」と対応させて表示されるものである。また、抑止コンテンツ93は、例えば、図示のように、表示される。

【0073】(ステップS117のアクセス監視・規制の決定)図22は、アクセス監視・規制処理についてのフローチャートである。まず、キャッシングモジュール42は、監視・規制前処理を行う(S117-1)。ここで、ステップS117-1の監視・規制前処理では、次の処理を行う。

①URIエントリを作成していなければ作成する。

②①を実行した場合は、URIハッシュに登録する。

③URIエントリ中のアクセスカウンタを1増やす。

【0074】つぎに、キャッシングモジュール42は、規制を行うか否かを判定する(S117-2)。ここで、ステップS117-2の規制判定処理では、①URIエントリのアクセスカウンタ、②URIエントリの規制実行フラグ、③装置に設定された規制開始接続数、④装置に設定された規制終了接続数を参照して、②の規制実行フラグが設定されていない場合、①>③ならばYesとし、そうでなければNoとする。一方、②の規制実行フラグが設定されている場合、①>④ならばYesとし、そうでなければNoとする。

【0075】キャッシングモジュール42は、ステップS117-2で規制を行う場合、規制メッセージを取得し(S117-3)、規制メッセージよりコンテンツを作成して(S117-4)、クライアントへコンテンツを転送する(S117-5)。一方、ステップS117-2で規制を行わない場合、キャッシングモジュール42は、Webサーバへ要求を転送し(S117-6)、Webサーバから応答を受信して(S117-7)、クライアントへ応答を転送する(S117-8)。

【0076】また、キャッシングモジュール42は、ステップS117-5、8の後、監視・規制後処理を行う(S117-9)。ここで、ステップS117-9の監視・規制後処理では、次の処理が行われる。

①URIエントリ中のアクセスカウンタを1減らす。

②アクセスカウンタが0ならば、URIハッシュから削除する。

③②で削除を実行した場合は、URIエントリを削除する。

【0077】図23は、規制メッセージ、規制コンテンツの説明図である。アクセス規制メッセージの作成方法についても、上述したアクセス抑止コンテンツの作成方法と同様である。規制メッセージ95は、例えば、「現在、混み合っているため後でご利用ください」と対応させて表示されるものである。また、規制コンテンツ97は、例えば、図示のように表示される。

【0078】なお、キャッシングモジュール42では、例えば、ステップS117-2において、以下のような処理を行い、規制を行うか否かの判断基準としてもよい。

A. URIエントリ新たなパラメータを追加する。

B. パラメータには、過去にWeb通信中継装置がサーバに当該URIの要求を送信したときの応答時間の平均を入れる。

C. アクセス数が1以上のときは、A. で追加したパラメータも考慮し、例えば、アクセス数がある程度存在し、通信に設定時間以上かかりそうな場合は、上述の規制開始接続数に満たなくても規制する。

【0079】（ステップS119の制御なしに決定）図24は、制御なし処理についてのフローチャートである。まず、キャッシングモジュール42は、Webサーバへ要求を転送し（S119-1）、Webサーバから応答を受信して（S119-2）、クライアントへ応答を転送する（S119-3）。ここで、ガベージコレクション（装置内資源の再利用）について説明すると、キャッシュ領域に空きが無くなると、データ格納が古かったり、アクセス回数が低い領域を消去して再利用するために、ガベージコレクション処理を行う。ガベージコレクション処理の際には、優先キャッシュエントリは対象外とし、さらに優先キャッシュフラグを参照して、優先キャッシュとして使用されているデータ領域も処理の対象外とする追加判定を行っている。

【0080】なお、通信システムがRFC2616に未定義のリクエストメソッドに対応しないときは、制御なし（S119、S211）を設置することなく発明の効果が得られる。また、通信システムがHEADメソッドをキャッシュとして処理しない場合は、ステップS101におけるHEADメソッドの処理を②ではなく③で処理することで対応できる。以下に、アクセス監視・規制に用いる規制開始接続数と規制終了接続数を動的に変更する場合について説明する。

【0081】図25は、拡張したURIエントリを示す図である。URIエントリを図25に示すように拡張することで、アクセス監視・規制に用いる規制開始接続数と規制終了接続数は、あらかじめ装置に設定した値だけでなく、動的に変更した値を利用できる。例えば、規制開始時刻と規制終了時刻により、次の規制開始接続数と規制終了接続数を定める。

【0082】図26は、規制開始接続数と規制終了接続数の動的な変更を示す説明図である。ここでは、規制開始接続数と規制終了接続数を動的に決定する場合の一例で、URI1とURI2に対する接続数の変化を示している。ここでは、URI1およびURI2とも、図25に示したURIエントリを利用しており、規制開始接続数と規制終了接続数にそれぞれ初期の接続数が設定されている。図26では、URI1とURI2が時刻t1にアクセス規制の制御が開始されている。この例では、規制開始を行うときにそれぞれのURIエントリに規制開始時刻t1を記憶する。規制が終了すると、t1と規制終了時刻t2およびt3を利用して、規制実施時間を求める。URI1の規制実施時間T1はt2-t1、URI2の規制実施時間T2はt3-t1となる。これを、通信システムが想定している標準規制実施時間Tと比較し、次の規制開始接続数と規制終了接続数を定める。この例では、T1の値がTより非常に小さいので、URI1の規制開始接続数と規制終了接続数を一定数増加している。また、T2は、Tとそれほど変わらないため、値を変更していない。

【0083】また、図25のように拡張したURIエントリでは、アクセス監視・規制を異なる方法で実施できる。Web通信ソフトウェア31内部のプロキシエンジン利用領域には、Web通信ソフトウェア31がWeb通信装置3-1～3-nに対して要求した際の平均応答時間が保存されている。さらにURIエントリには、URIごとの平均応答時間が記憶されている。URIに対して一定数の要求がある場合、Web通信ソフトウェア31全体の平均応答時間とURIごとの平均応答時間を比較し、URIごとの平均応答時間の応答時間が著しく大きい場合は、アクセス規制処理を行うことで輻輳を回避できる。

【0084】

【発明の効果】本発明を利用することにより、通信システム提供者は、従来よりも信頼性のある通信システムをユーザに提供することができる。また、ユーザは、従来よりも少ない待ち時間で応答の得られる通信システムを利用できる。

【図面の簡単な説明】

【図1】通信システムの構成図。

【図2】Web通信中継装置の構成図。

【図3】大規模Web通信中継装置におけるデータの送受信についての説明図。

【図4】キャッシュヒット時の説明図及びシーケンス図。

【図5】キャッシュミスヒット時の説明図及びシーケンス図。

【図6】輻輳時の説明図及びシーケンス図。

【図7】アクセス抑止対象コンテンツが要求された場合の説明図及びシーケンス図。

【図8】Web通信中継ソフトウェア31の内部構成についての説明図。

【図9】URIハッシュとURIエントリによるURI管理についての説明図。

【図10】共用メモリ55内のデータ構造についての説明図（1）。

【図11】共用メモリ55内のデータ構造についての説明図（2）。

【図12】キャッシング及び輻輳制御処理について基本的な処理を示すフローチャート。

【図13】キャッシュチェックおよび輻輳制御処理判定についてのフローチャート。

【図14】クライアントのHTTP要求についての説明図。

【図15】URIの加工についての説明図。

【図16】通常キャッシュ処理についてのフローチャート。

【図17】Web通信装置のHTTP応答の説明図。

【図18】Webコンテンツデータの管理についての説明図。

【図19】優先キャッシュ処理についてのフローチャート。

【図20】アクセス抑止処理についてのフローチャート。

【図21】抑止メッセージ、抑止コンテンツの説明図。

【図22】アクセス監視・規制処理についてのフローチャート。

【図23】規制メッセージ、規制コンテンツの説明図。

【図24】制御なし処理についてのフローチャート。

【図25】拡張したURIエントリを示す図。

【図26】規制開始接続数と規制終了接続数の動的な変更を示す説明図。

【符号の説明】

1-1~1-i クライアント端末

2 大規模Web通信中継装置

3-1~3-j Web通信装置

21 データ中継装置

22 認証装置

23 システム管理装置

24 レイヤ7負荷分散装置

25-1~25-n Web通信中継装置

26 アクセス規制装置

31 Web通信中継ソフトウェア

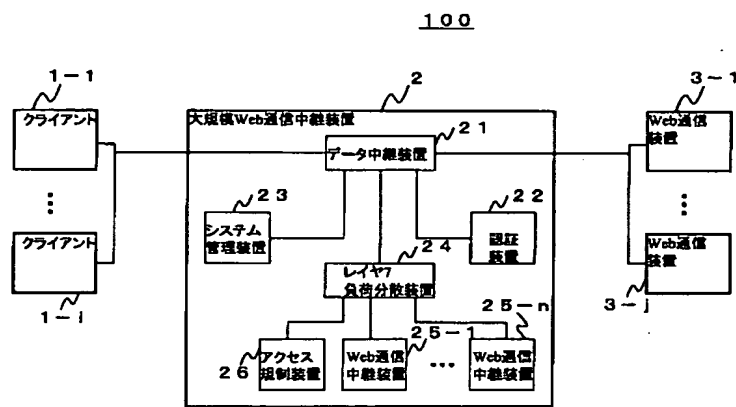
41 HTTPモジュール

42 キャッシングモジュール

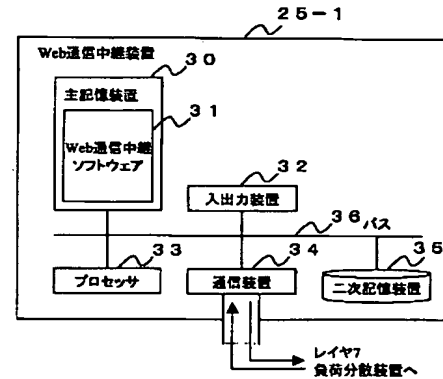
43 プロキシエンジン

100 通信システム

【図1】

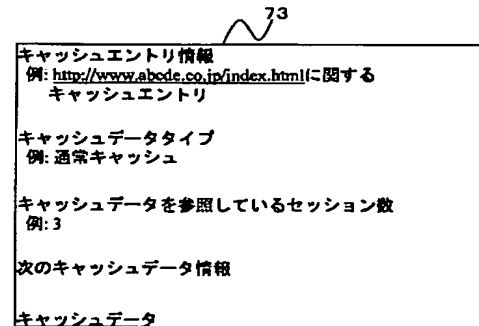
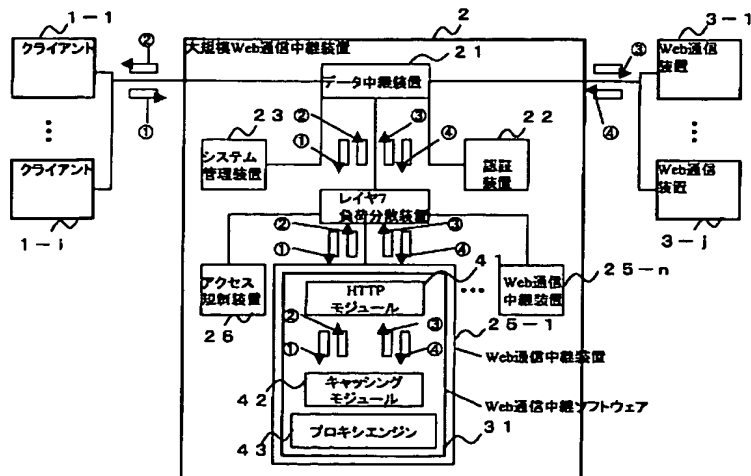


【図2】

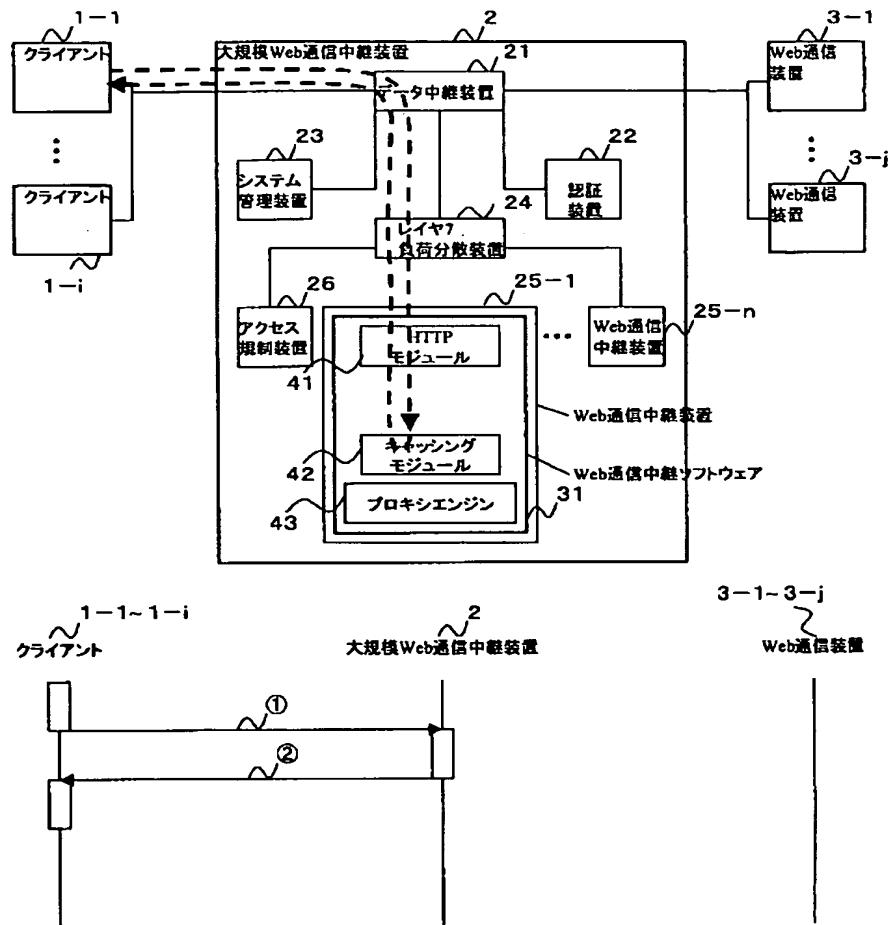


【図11】

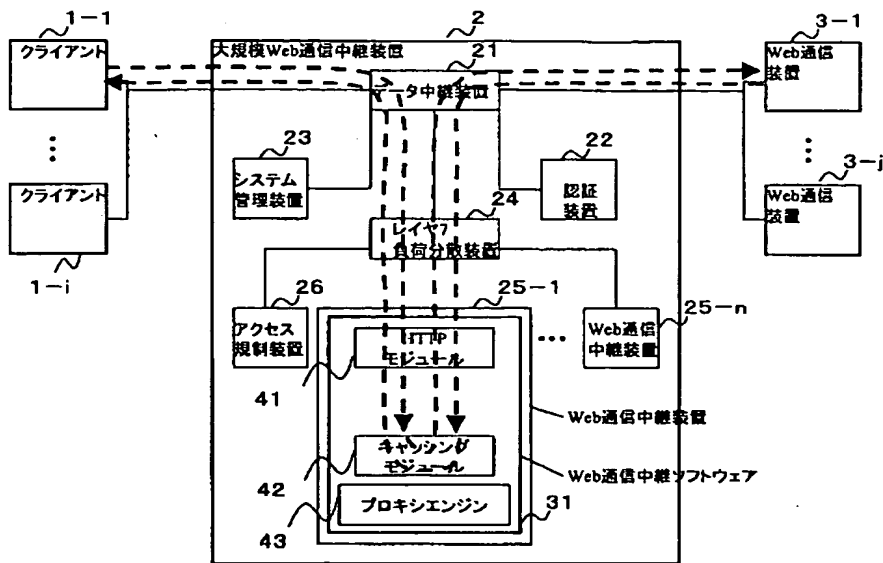
【図3】



【図4】



【図5】



【図15】

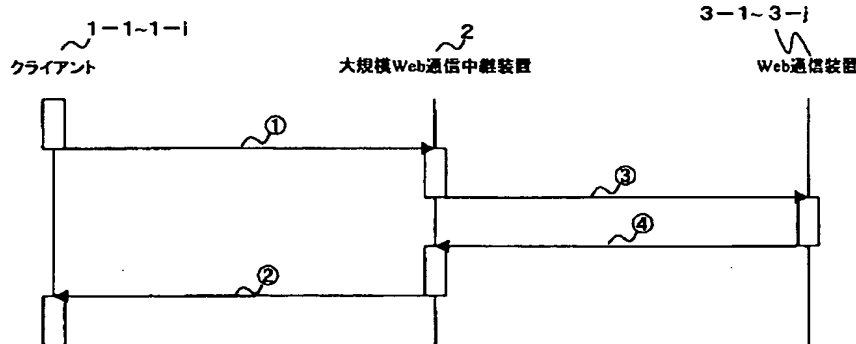
スキーム パス
<http://www.abcd.co.jp/index.html>

URI

<http://www.abcd.co.jp/cgi?al=arg1&a2=arg2>

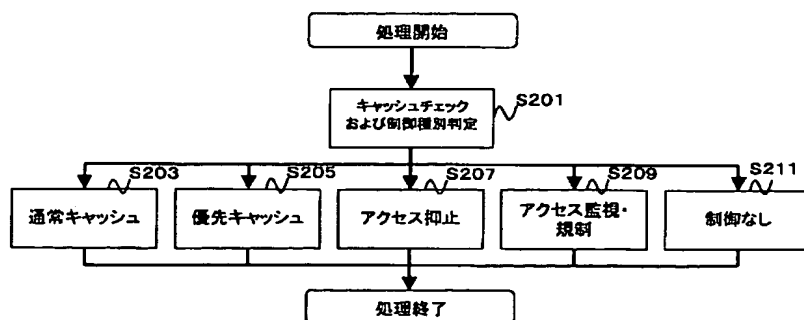
<http://www.abcd.co.jp/cgi>

URIの加工



【図12】

【図14】



リクエストメソッド URI 80

GET <http://www.abcd.co.jp/> HTTP/1.1

Proxy-Connection: Keep-Alive

User-Agent: Web Browser version 1 [en]

Host: www.abcd.co.jp

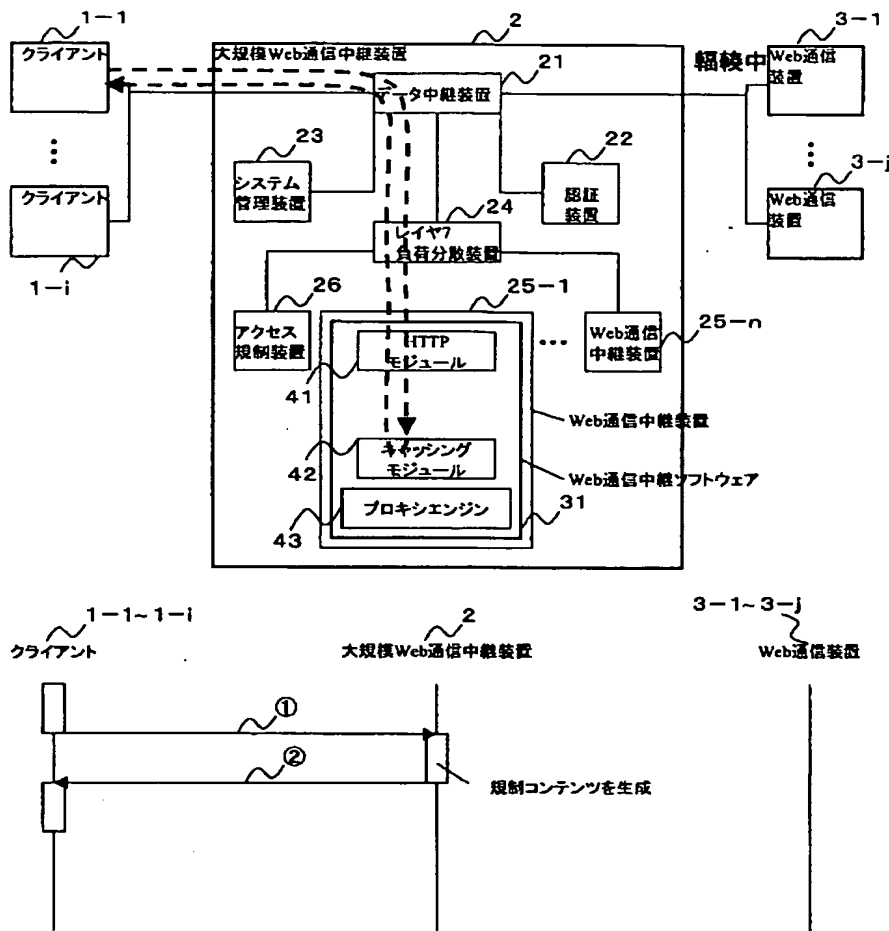
Accept: image/gif, image/x-bitmap, image/jpeg, image/png, */*

Accept-Encoding: gzip

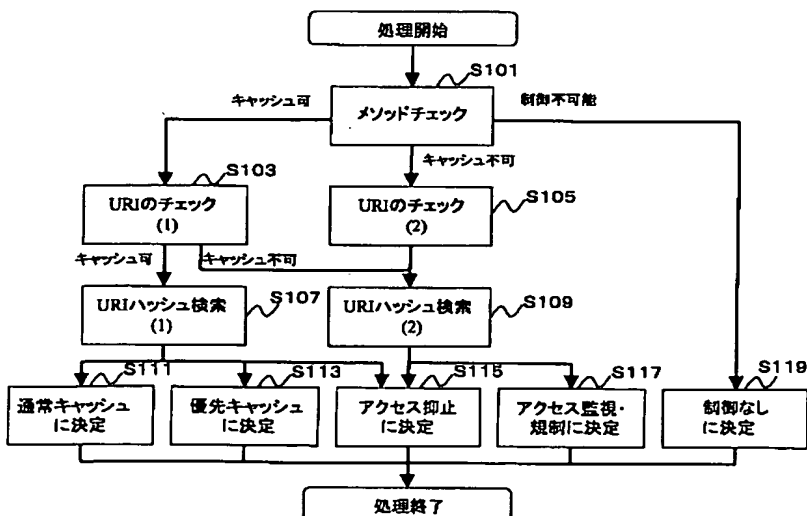
Accept-Language: en

Accept-Charset: iso-8859-1,*,utf-8

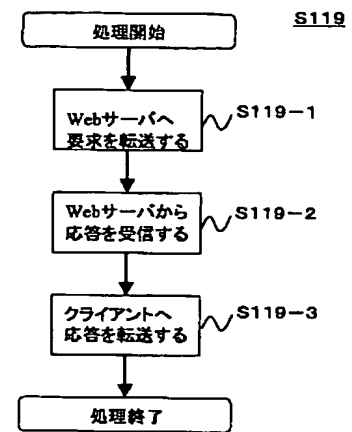
【図6】



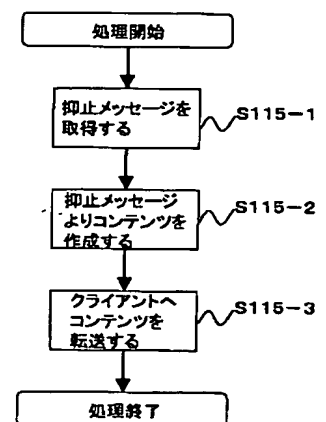
【図13】



【図24】

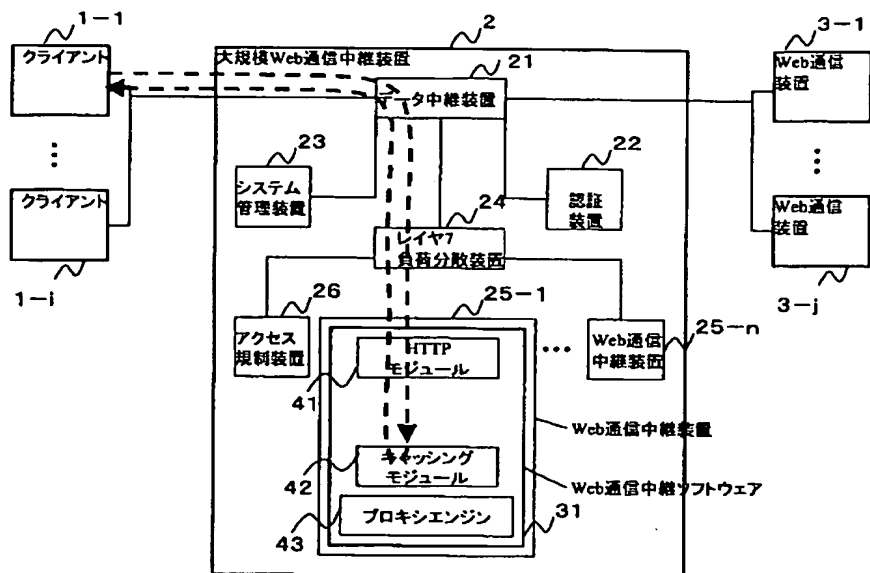


【図20】



【図7】

【図25】



URI
例: <http://www.abode.co.jp/index.html>

制御タイプ
例: 通常キャッシュ

アクセスカウンタ
例: 3 アクセス

規制実行フラグ
例: 規制なし

キャッシュエントリ情報

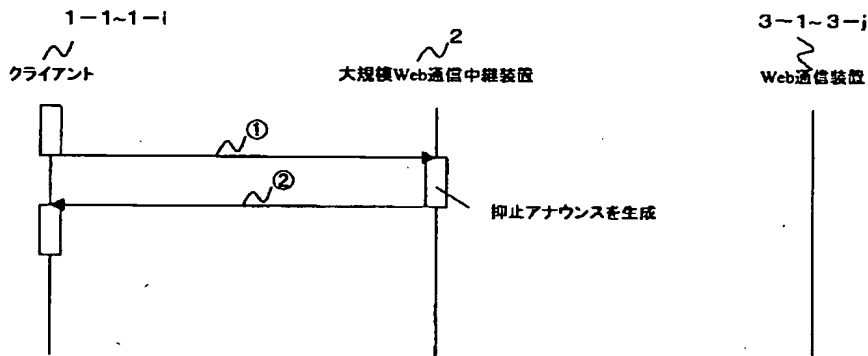
累計接続数
例: 17回

平均処理時間
例: 500ms

規制開始接続数
例: 64アクセス

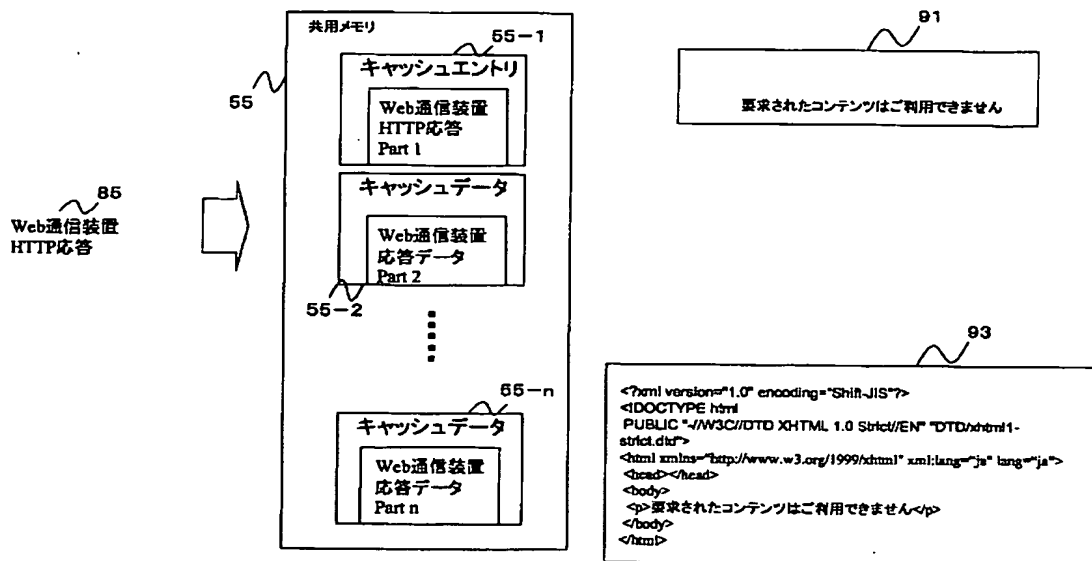
規制終了接続数
例: 32アクセス

規制開始時刻
例: 2001年7月1日0時0分0秒

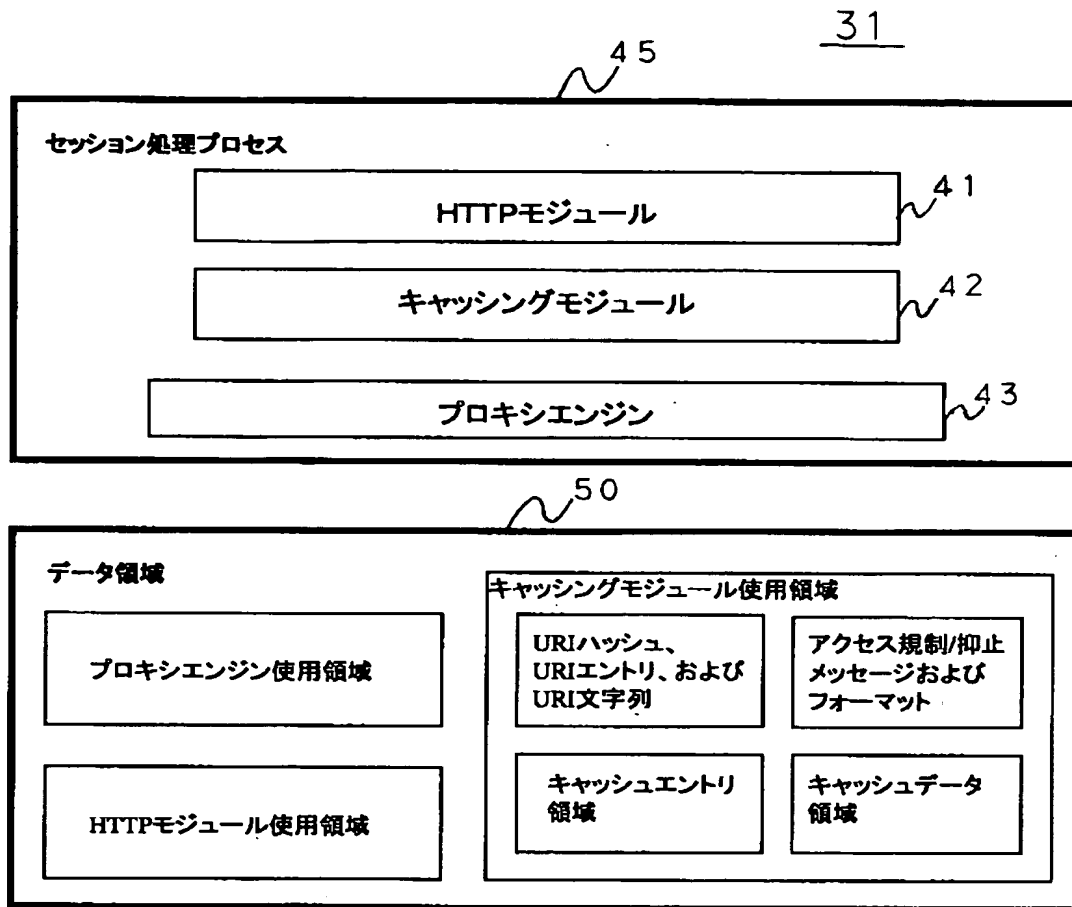


【図18】

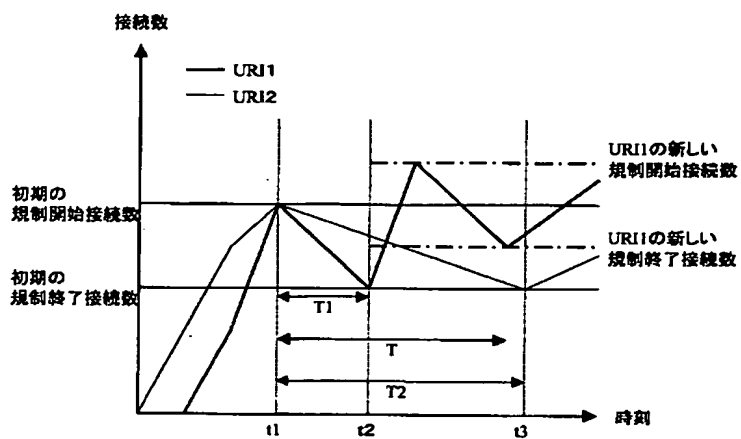
【図21】



【図8】



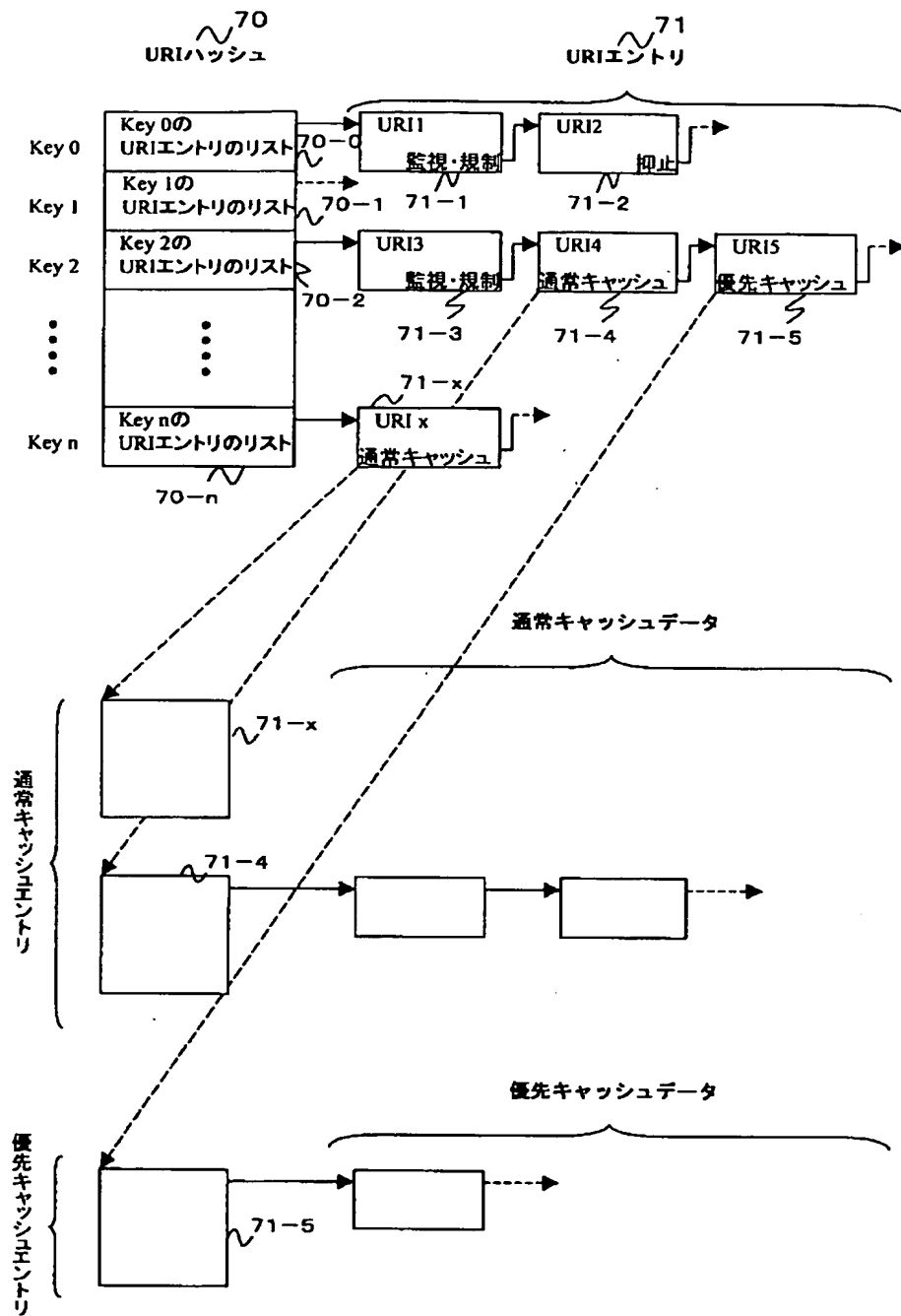
【図26】



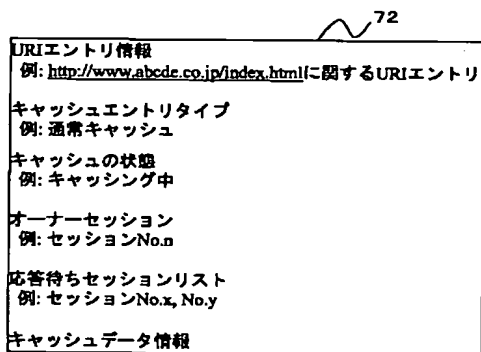
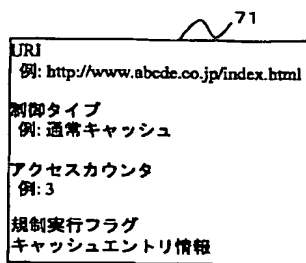
$$T1 \ll T$$

$$T2 > T$$

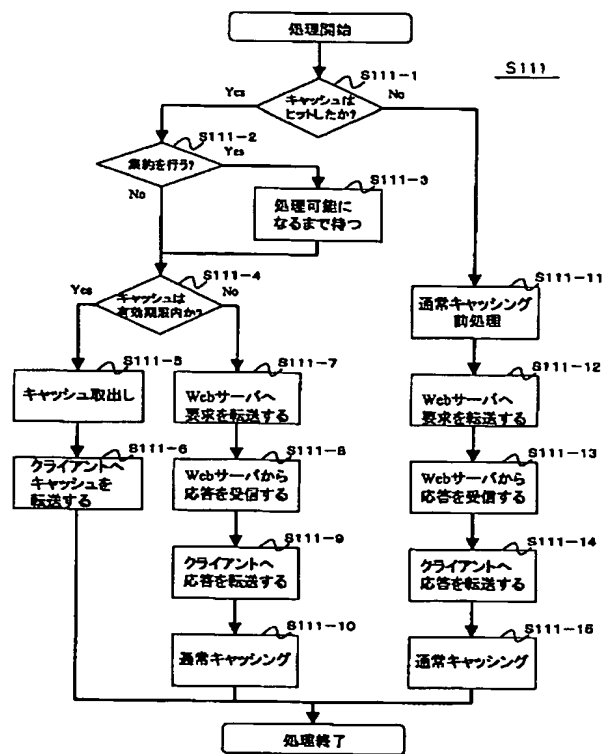
【図9】



【図10】



【図16】



【図17】

85

HTTP/1.1 200 OK

Date: Thu, 08 Feb 2001 13:31:07 GMT

Server: www server version 1.0

Last-Modified: Thu, 30 Nov 2000 07:04:44 GMT

Accept-Ranges: bytes

Content-Length: xxx

Connection: close

Content-Type: text/html

<html>

<head>

<title>Web Page</title>

</head>

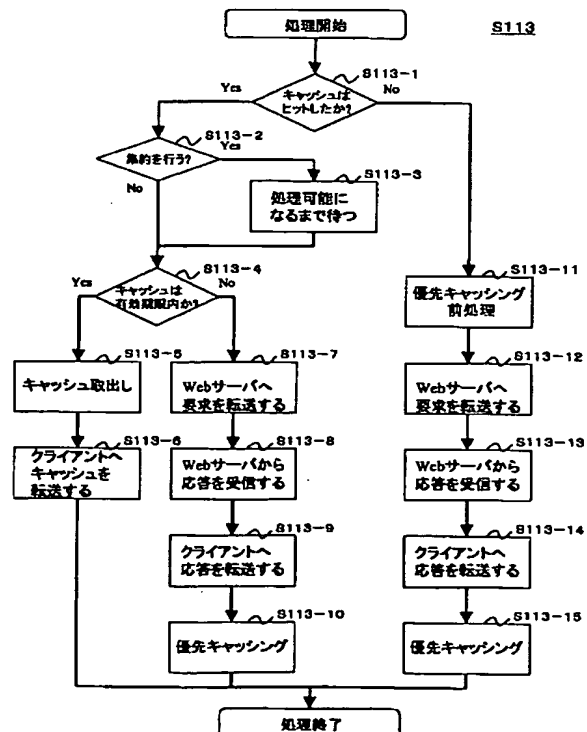
<body>

...

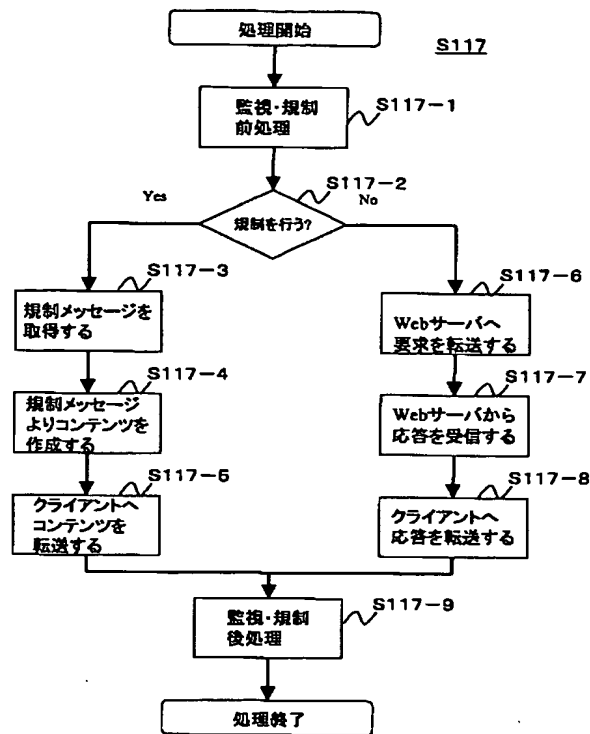
</body>

</html>

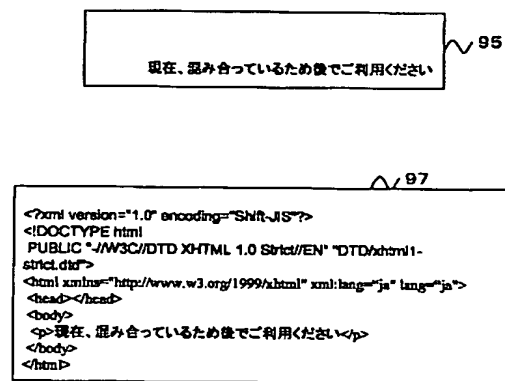
【図19】



【図22】



【図23】



フロントページの続き

(72) 発明者 横田 大輔
東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内

(72) 発明者 高橋 泰弘
神奈川県横浜市戸塚区戸塚町216番地 株
式会社日立製作所ネットワークプラットフ
ォーム事業部内

(72) 発明者 野田 文雄
東京都国分寺市東恋ヶ窪一丁目280番地
株式会社日立製作所中央研究所内

(72) 発明者 竹島 由晃
神奈川県川崎市麻生区王禅寺1099番地 株
式会社日立製作所システム開発研究所内

Fターム(参考) 5B082 FA12
5B089 GA19 GB01 JA21 KA06 KA07
KC18 KD02 MA03